Probability of Precipitation (PoP) Forecasts

The National Weather Service has been producing PoP forecasts for the past fifty years. This entire time, we have applied statistics to produce and evaluate these forecasts.

In the forecast process, first the initial data (temperature, pressure, wind, humidity) are collected to describe the state of the weather before the forecast is produced. Then, the numerical models are run to apply the laws of motion and thermodynamics to the forecast process. The model provides forecast data (temperature, pressure, wind, and humidity). These forecasts are entered into statistical models, which apply the memory of recent weather (the last few years) to the model output. The statistical models always produce a map of the United States with PoP forecasts for every 12-hour period during the next seven days, e.g., 8 am to 8 pm EDT. The forecaster has the choice of selecting which model output to apply to the current forecast. Several models or model ensembles may be combined to produce an initial version of the forecast. Often, the forecasts need to be adjusted for local effects, caused by coastlines, lake shores, or mountains.

Here is what we are forecasting as a 12-hour PoP for Prince Georges County: the probability that measurable precipitation (0.01 inch) will fall at any given point in that area during the prescribed 12-hour forecast period. The best way to subjectively describe "measurable" rain is that it is the amount of rain that it takes to entirely wet the ground and begin to produce puddles over roads or sidewalks. Brief sprinkles that barely wet the entire ground do not produce enough rain to qualify. These 'light sprinkles' tend to produce a trace of rain, i.e., more than zero rainfall, but less than 0.01 inch.

Forecast Verification. This is when we look at a lot of points in the country and evaluate the quality of the forecast at each point by comparing the weather forecast to a weather observation taken at the surface of the earth. Sometimes, estimates of those ground observations can be used in sparsely populated areas, from which no weather observations are available. Every half day (12-hour) period, the forecast is compared to the rain gage observation at each data point, and the following Brier score is calculated for a set of forecasts. The sigma notation used in this equation is explained in a separate hand-out.

$$B = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2$$

where,

i = a counter for each forecast in the calculation; i runs from 1 to N, where N is the highest i value.

f_i = the value of each PoP forecast in the calculation, a value from zero to 100 exists for each i.

o_i = the ith rain gage observation matched in space and time to the ith forecast value at the appropriate forecast projection. If the precipitation for a particular 12-hour period was measurable, then o_i equals 100. If no measurable precipitation occurred, then o_i equals zero.

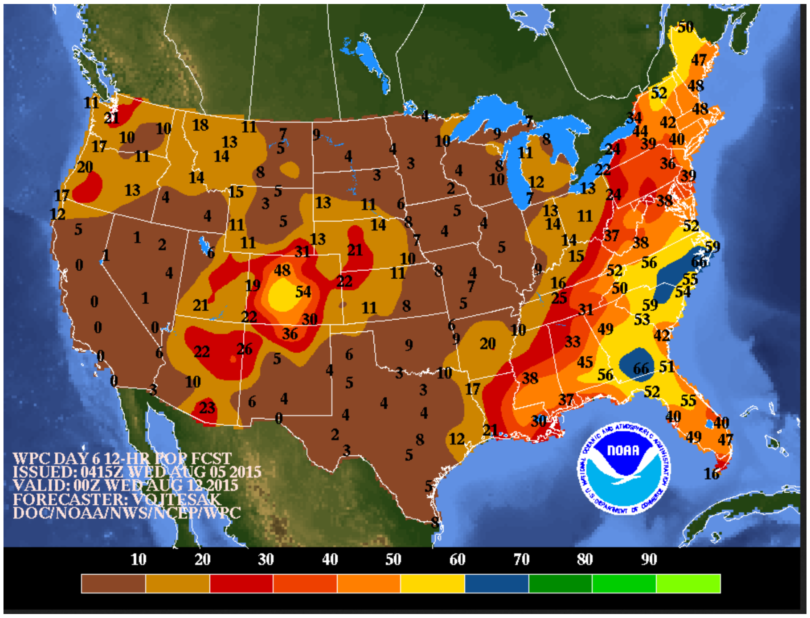
Example.

If the PoP at a given point in space at a given forecast projection was 70%, and measurable rainfall was observed during that 12-hour period, then the contribution to the Brier score from this single event, the 74th member in a dataset with 300 members, is:

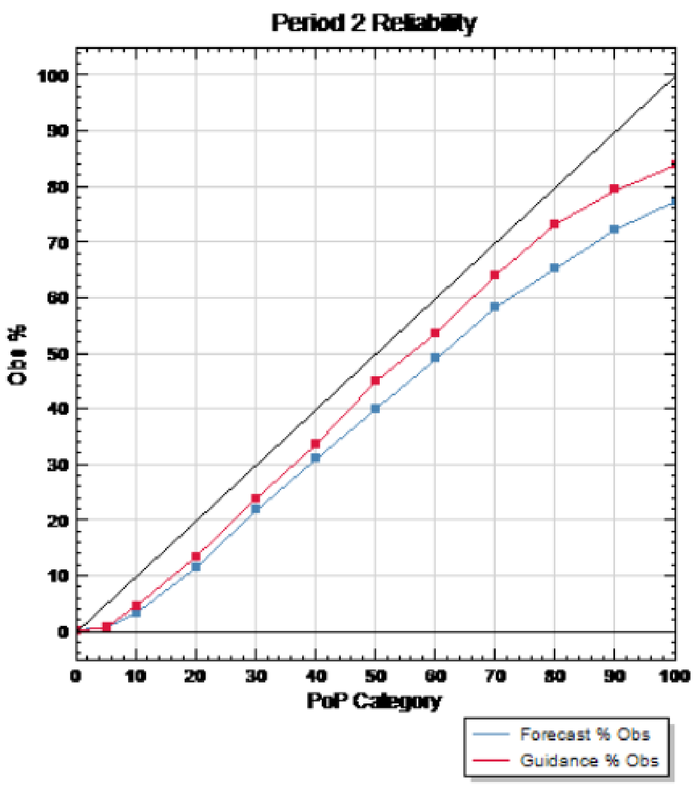
$$B = \frac{1}{1} \sum_{i=74}^{74} (0.7 - 1.0)^2$$

The **Reliability Diagram** provides a "long-term" plot of PoP forecast bias. First, each PoP value, expressed as a percentage from zero to 100, is rounded to the closest of one of the following values: 0, 5, 10, 20, 30, 40, … , 80, 90, 100. The matching observation in time and space, categorized as 'measurable precipitation' or 'no measurable precipitation,' is compared to each forecast. The frequency of the number of observed measurable precipitation events that occurred when each of the twelve rounded PoP values was predicted is recorded and converted to the percentage of all forecast/observation sets included in the plot. The observed percentage of measurable precipitation occurrences when each of the twelve rounded PoP values was predicted is plotted as the ordinate for each point on the reliability diagram; the rounded PoP value is the abscissa for each point. The twelve plotted points are connected with a blue or red line (the red line represents the model guidance provided to the forecaster). The portions of the blue or red curve above the black diagonal line represent the under-forecasted PoPs, and the portions below the black diagonal line represent the over-forecasted PoPs.

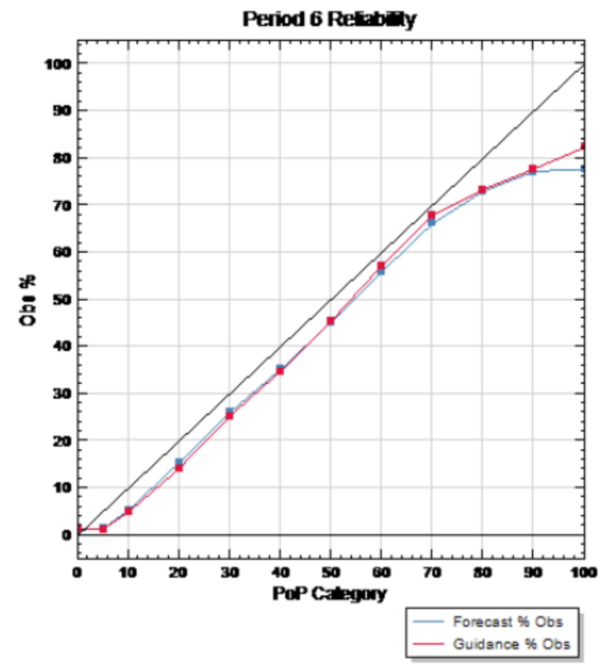
In the long run, a fully reliable (unbiased) set of rounded 10 percent PoP forecasts occurs exactly 10 percent of the time that the rounded PoP is 10 percent; a fully reliable (unbiased) set of rounded 20 percent PoP forecasts occurs exactly 20 percent of the time that the rounded PoP is 20 percent, etc., all the way to 100 percent. Ideally, no measurable precipitation events occur when the rounded PoP forecast is zero, and every 100 percent rounded PoP forecast has an observed measurable precipitation event.



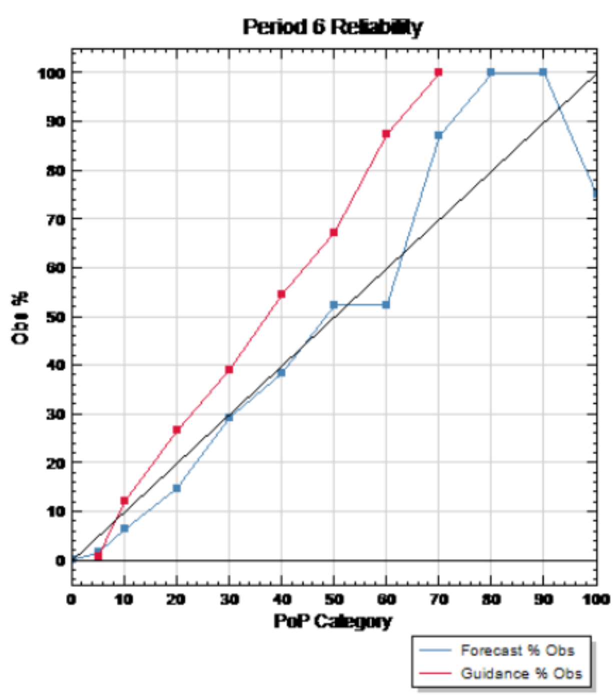
Sample six-day Probability of Precipitation forecast, issued at 12:15 am EDT, Wednesday, August 5, 2015, and valid from 8 am to 8 pm EDT, Tuesday, August 11, 2015.



PoP reliability curve for 12-24 hour forecasts, entire United States, FY11. Both curves (forecaster and model) are entirely below the black diagonal line, which means PoP values were over-forecasted.



PoP reliability curve for 60-72 hour forecasts, entire United States, FY11. Both curves (forecaster and model) are entirely below the black diagonal line, which means PoP values were over-forecasted.



PoP reliability curve for 60-72 hour forecasts, Arizona, FY11. In this desert state, the model (red line) under-forecasted the PoPs at the 20 thru 70 PoP intervals, but the forecasters (blue) largely corrected that bias. The model did not predict 80 thru 100 PoP intervals; the forecasters demonstrated an under-forecasting bias at these intervals.